Coll. Antropol. **36** (2012) 1: 31–37
Original scientific paper

Data Analysis Strategies for Reducing the Influence of the Bias in Cross-Cultural Research

Joško Sindik

Institute for Anthropological Research, Zagreb, Croatia

A B S T R A C T

In cross-cultural research, researchers have to adjust the constructs and associated measurement instruments that have been developed in one culture and then imported for use in another culture. Importing concepts from other cultures is often simply reduced to language adjustment of the content in the items of the measurement instruments that define a certain (psychological) construct. In the context of cross-cultural research, test bias can be defined as a generic term for all nuisance factors that threaten the validity of cross-cultural comparisons. Bias can be an indicator that instrument scores based on the same items measure different traits and characteristics across different cultural groups. To reduce construct, method and item bias,the researcher can consider these strategies: (1) simply comparing average results in certain measuring instruments; (2) comparing only the reliability of certain dimensions of the measurement instruments, applied to the »target« and »source« samples of participants, i.e. from different cultures; (3) comparing the »framed« factor structure (fixed number of factors) of the measurement instruments, applied to the samples from the »target« and »source« cultures, using explorative factor analysis strategy on separate samples; (4) comparing the complete constructs (»unframed« factor analysis, i.e. unlimited number of factors) in relation to their best psychometric properties and the possibility of interpreting (best suited to certain cultures, applying explorative strategy of factor analysis); or (5) checking the similarity of the constructs in the samples from different cultures (using structural equation modeling approach). Each approach has its advantages and disadvantages. The advantages and lacks of each approach are discussed.

Key words: *importing constructs, latent dimensions, measuring instruments*

Introduction

Due to cultural differences, it is impossibe to simply copy strategic economic and political solutions, as well as scientific construct(s), especially in psychology.

However, much research contrasts self-reported personality traits across cultures. These research results could be a consequence of a using a specific metric approach or a consequence of qualitative differences in the theoretical concepts, according to the heritage in different cultures. In other words, in order to conduct cross--cultural research with good validity we have to adjust the construct(s) and associated measurement instruments which have been developed in one culture and then imported for use in another culture. In point of fact, though, importing concepts from other cultures is often simplified reduced on

language adjustment for the measurement instruments that define a particular construct. In multigroup cross-cultural comparisons, typically the measuring instruments have been translated from the language of the »source« country in which it was developed and normed into the language of the »target« country in which it is to be used (hereinafter the terms country and cultures will be used as synonyms)[1]. It is typically assumed that the instrument of measurement is operating in exactly the same way across cultures, and that the underlying construct(s) has the same theoretical structure and psychological meaning across the groups of interest. But very frequently we can find the arguments for the instrument nonequivalence within a cross-cultural context. That fact could be considered as some aspect of test bias in the operation of the measurement instrument. Test bias can be described as a tendency of a test to systematically over- or underestimate the true scores of individuals to whom that test is administered, or those who are members of particular groups (ethnic minorities, sexes, *etc.*)[2]. More simply, it means that test bias is generally interpreted as

Received for publication October 17, 2011

an indicator that the items are differentially valid across selected groups[1]. In the context of cross-cultural research, van de Vijver and Leung[3] define »bias« as a generic term for all nuisance factors threatening the validity of cross-cultural comparisons. Test bias can be an indication that test scores based on the same items measure different traits and characteristics for across groups.

Three Types of the Test Bias

Vijver and Tanzer[4] consider the issue of bias as not necessarily related to the intrinsic properties of an assessment instrument per se. The bias can reflect the characteristics of the participants from each cultural group. The statements regarding bias always refer to the use of an instrument within the framework of particular applications of cross-cultural comparisons. Authors differentiated three primary sources (types) of bias: (a) the construct of interest (construct bias), (b) the methodological procedure (method bias), and (c) the item content (item bias). Construct bias refers to the fact that measured construct have some degree of differential meaning cross the different cultural groups[1]. There are three potential sources of differential meanings: (a) the behaviors chosen as indicators of a construct can be differentially appropriate across cultural groups (for example, behaviors that describe the concept of being a »good« son or daughter, have a different meaning in Asian and in Western societies); (b) the extent to which all relevant dimensions of the construct are included in the formulation of the item content varies across groups, even if the item content is the same (for example, the facet of the emotional self concept from the »source« population may be totally meaningless or irrelevant in »target« cultural groups in comparative study); (c) the sampling of behaviors representative for the constructs that being measured may be inadequate for a particular cultural group (for example, in cultures with closely linked extended families, much broader range of social interactive behaviors are needed for defining the same social concept from more individualistic cultures)[1]. Method bias refers to the three specific aspects of the methodological strategy employed in testing for equivalence across the cultural groups. (a) Sample bias refers to the comparability of samples on phenomena other than the target factors under study (for example, if we compare the knowledge of grammar in the same foreign languages in two different cultures, it is possible that the curriculum of study in one culture might emphasize acquired all aspects of language learning, while only grammar and writing ability is primarily important in other culture). (b) Instrument bias relates to the differential responses by comparative groups, to the structured format of the assessment instrument, with two aspects: (b1) stimulus familiarity refers to the possibility that the type of stimulus response like a Likert scale may be unfamiliar to some cultural groups; (b2) patterns of response refers on two phenomenon: response style – consistently selecting one of the two extreme scale points (high, low)[5,6,7], independent of the item content; response set – selecting scale points in a way as to convey a favorable impression of oneself (e.g., social desirable responses). (c) Administration bias describes the discrepancy in the administration of an instrument to the participants of the comparative groups (for example, one group may have the opportunity for to completion a set of practice items, while the other group does not). Item bias refers to distortions at the item level, mainly influenced with eliciting a differential meaning of their content across cultural groups. Members of culturally-different groups have a different heritage and live in a diversity of sociocultural contexts that include the family, the school, the peer group, and society at large[1].

A Proposal for Data Analysis Strategies for Reducing Test Bias in Research

A review of the literature in cross-cultural psychology, as well as in psychology in general, or in the other social sciences, indicates that considerable research uses measurement instruments imported from different countries while not sufficiently accounting for cultural adjustments in the instruments. Through a literature review of those studies providing a more detailed descriptions of instruments' cultural adjustment, we can differentiate a few types of research approaches used in analyzing cultural differences. This review takes into account reported data analysis methods and strategies used in reducing the influence of the bias in the cross-cultural research. While some of these methods are very complex and sophisticated, this short review focuses on five types of those approaches.

(a) »Simply« comparing the results in certain measurement instruments in different countries. Researchers usually explore the factor structure of the instrument, but for all the results taken together (in »target« and »source« country). The next step is the definition of the scores on certain instrument dimensions and finding the differences between different cultures using ANOVA, MANOVA or univariate procedures (t-test, Mann Whitney U-test, etc.). In this type of approach, the authors do not neglect the possibility of the influence of the cultural bias, but do not choose very restrictive methods to avoid it. Such approaches have multiple deficiencies. Mainly, they underestimate the influence of construct bias (there is no factor analysis conducted on separate samples of participants from different cultures), method bias and item bias. The advantage of such an approach is its simplicity. On the other hand, it could be a quite good method for research among members of similar cultures. Here are some examples, which try to go beyond »simple comparison«. For example, back translation of the instruments and correcting the meaning of the estimation scales constitute some of the more sophisticated methods, apart from making a simple comparison. Gartstein et al.[8] investigated the early development of temperament across four cultures: Japan, United States of America (U.S), Poland and Russia, through a cross-sectional design. They compared four cultures that vary on indi-

vidualistic and collectivistic value systems. Parents responded to the Infant Behavior Questionnaire-Revised. The MANOVA approach was utilized initially with the three overarching temperament factors as dependent variables in order to limit the overall number of analyses. The authors evaluated the main effects of culture, age and gender, as same as their interactions, using common factors, in all cultures together. The primary aim of the study conducted by Chen et al.[9] was to examine cultural differences in Chinese and American adolescents and parents' perceptions and evaluations of adolescent misconduct behaviors. Each participant made a list of adolescent misconduct behaviors and rated each misconduct behavior, as to the degree of misconduct. Significant cultural differences were found in a number of adolescent misconduct behaviors. For example, Americans generated more misconduct behaviors in offenses with weapons and drug use than did Chinese. These cultural differences were further complicated by an interaction between the role of culture and that of family.

Christie et al.[10] used a more complex methodology for data analysis than a simple comparison, trying to avoid some aspects of method bias. They conducted a study of ethical attitudes of business managers in India, Korea and the United States. This study examine the relationship between five cultural dimensions (individualism, power distance, uncertainty avoidance, masculinity and long-term orientation) and business managers' ethical attitudes. The results indicate that national culture (cultural dimensions of individualism and power distance) has a strong influence on business managers' ethical attitudes. However, authors did not neglect the need for an instrument measuring uniformly across cultures and without cultural bias, is imperative for the validity of any cross-cultural study. The most highly regarded solution to avoid cultural bias suggested by researchers is de-centering, a process in which researchers from different cultures are involved in developing research questions[11–13]. The instruments were reviewed by a panel of experts in linguistics from three cultures (India, Korea and the United States) for its clarity and domain appropriateness, to ensure its content validity. The results of the pilot study showed that in general the ethical attitudes of respondents differed in accordance with theoretical expectations basically ensuring construct validity[14]. The instruments showed satisfactory reliability coefficients, but the analysis was performed for all the participants (from all the cultures together). One specific method (or set of methods) is suggested by Heine et al.[15], which is based on analyzing a few ways for avoiding bias in cross-cultural comparisons of subjective Likert scales. One challenge is the reference-group effect (RGE)[15,16], the tendency for people to respond to subjective self-report items by comparing themselves with implicit standards from their culture[15]. Their strategies for the cross-cultural comparisons that protect against reference-group effects were few. However, they believe that the that the most successful approach is likely one that combines many of these and uses multiple methodologies[17]. Their methods are mainly based on considering the combination of physiological, behavioral, demographic measures, peer-reports and self-reports, as well as the instruments that measure the same or very similar construct. That is, in fact, the method based on partial correlation coefficients in defining the validity of the cross-cultural studies[15].

The critical assumption in this approach is that the translated instrument is equivalent to the original instrument with respect to its reliability and validity, as well as its norms[1,18]. However, these assumptions are very often absolutely wrong[3,19], because reliability and validity need to be tested in each multicultural sample. In the comparison between groups different in their culture, the sources of possible bias are numerous, in the variety of multicultural groups and their societal structures[1]. Revision of a measuring instrument developed in one country for use in another, regardless of language itself, ultimately makes a different instrument. So, perfect instrument translation is not the guarantee for its »immunity« from bias. Statistically rigorous testing is needed in order to equalize psychometrical characteristics of its original (»source«) and »target« version.

(b) Comparing only the reliability of certain dimensions of the measurement instruments, can be applied on the samples from the »target« and »source« samples of participants (from different cultures). We compare the reliability of each dimension of the original certain measurement instruments (constructed in the »source« country), this time both in »source« and »target« countries, and then find differences between members of different cultures. In fact, we using this approach, we »automatically« use the same set of items that define some instrument dimensions in the »source« country as hypothetically valid for the »target« countries and check if the same items are reliable for the defining the »source« dimensions, for the each sample of the »target« countries separately. We can use the items that define a certain construct's dimensions in a »source« country, assuming that the same items define the dimensions in »target« countries metrically valid (without performing factor analysis of the whole space of items in »target« countries). Then, we can find out that the reliability of some dimensions in »target« countries is unsatisfactory (for example, below 0.50 expressed with a Cronbach's alpha coefficient): the possible reason can be existing cross-cultural differences. In other words, sometimes only the reliability coefficients could reflect cross-cultural differences. Nevertheless, it's the one important step forward in making cross-cultural comparisons, as contrasted with the first type of research approach, also has multiple deficiencies. This approach provides helpful information about the influence of c), method bias (one aspect of instrument bias) and item bias, but doesn't offer much of a practical solution to the question being addressed here. Considering reliability of the dimensions in samples of the »target« countries, we can take a few actions. At first, we can simply describe the reliability comparison between the dimensions in the samples from different cultures and give the directions for future research. Sec-

ondly, we can avoid comparisons between unreliable dimensions. In cases when the researcher find that some dimension in any culture is unreliable, we can compare the remaining reliable dimensions (except unreliable dimensions). Thirdly, we can redefine the unreliable dimension (in any sample) in such way that we omit the items that are the most unreliable (in each of the samples), until certain dimension gain satisfactory reliability. Then we can compare such reliability-modified dimensions in different cultures. The advantage of this approach is itssimplicity. Of course, it could be a quite good method for researching members of similar cultures. Here are some examples. Barrett, Sonderegger and Sonderegger[20] in their study examined whether young migrants, differentiated by cultural background, vary in their experience of cultural adjustment, emotional distress, levels of self-esteem, and coping ability. Students were recruited at random from public schools in South East Queensland, differentiated by cultural origin (former-Yugoslavian, Chinese, Mixed-culture, and Australian) and school level (primary and high school). Students completed measures of cultural adjustment, anxiety and trauma measures, self esteem and coping ability. This study reveals information on how culturally diverse migrants acculturate, the type and severity of symptoms they experience, and their capacity to cope in stressful situations. Participants' scores on the psychometric inventories were analyzed using a total of three ANOVAs and five MANOVAs. To ensure comparability between cultural groups, each self-report measure was analyzed for internal consistency, separately in each cultural group. Reliability coefficients were obtained for each assessment inventory among the different cultural groups. Pearson's correlations were calculated among the culturally-diverse self reports to determine whether these constructs relate in the same way. In another study conducted by Barrett et al. using the Reid Integrity scale (for predicting counterproductivity), with appropriate language translations, the authors examined samples in three countries, in Argentina, Mexico and South Africa[21]. The samples consisted of job applicants and current employees who completed the scale, while the supervisors provided performance evaluations for the majority of employees on the dimensions of counterproductive behavior, general work performance, social interaction and positive employee traits. Comparisons of mean scores and reliability coefficients indicated comparable responses to the scale across cultures and with US samples. The Reid Integrity scale Inventory was assessed within each sample and was found to be highly reliable. For Argentina, Cronbach's $\alpha = 0.81$, for Mexico, $\alpha = 0.78$ and for South Africa, $\alpha = 0.79$ (reliability coefficient obtained in a »source« country USA was $\alpha = 0.83$). They concluded that the instrument was appropriate for cross-cultural research. However, perfect instrument translation and reliability analysis is not the guarantee for its »immunity« against bias. More rigorous statistical testing is needed in order to equalize psychometrical characteristics of its original (»source«) and »target« version.

(c) Comparing »framed« factor structure of the measurement instruments, applied to the samples from the »target« and »source« cultures, using explorative factor analysis strategy on separate samples (from different cultures): Analyzing the factor structure of the space of measurement instruments, applied both in the »target« and »source« country (previously constructed in a »source« country), can be a more sophisticated method for avoiding test bias. We can apply the strategy of defining the number of factors in advance (the same number in »target« as it is in a »source« country). In fact, we perform the simple approach with explorative factor analysis techniques, with a given number of factors: we try to assess whether or not the factor structure and factor saturation in the »target« country is compatible with those in a »source« country. After factor rotation, we might find that factor structure is exactly the same, i.e. that the same items saturate the same factors; in this case we can perform simple comparison. The lack of this procedure can be in the same factor structure, but different factor loadings, so we have to be sustained in the results' interpretation). However, if we find that the factor structure is not compatible in a »source« and »target« country, we can remove from the analysis all the items in the same factors that are not compatible in two countries. In short, we find the intersection of the sets of items, check their reliability (in both countries separately), and then define factor scores for comparing the results between countries. The lack of this procedure can result in loosing some factors, because it's possible that too many items can be removed from the analysis. The third (and perhaps the most unpopular method among all the methods for comparing data from two or more cultures), is to compare the scores between the same »factors« in different cultures. We can simply summarize the scores for the same items that belong to certain factors (simple linear combinations) as the factors are the same in a »source« and »target« country, if the factors have approximately similar structure. When using this approach, the researcher needs to clarify its limitation. In the situation without defining number of factors in advance, a researcher has to compare the factor structure separately for each sample, but the analysis of the differences can be only descriptive. Richter et al.[22] in their study conducted a cross-cultural comparison of personality traits between individuals from two very different cultures and refugees who resettled several years before from one place to the other. They analyze the samples of Swedish individuals from the normal population and Iranian refugees in Sweden, as well as domestic Iranians from Tehran, with a questionnaire for assessment of their temperament and character. Principle axis factor analysis with Varimax rotation with Kaiser normalization was calculated for the Temperament and Character subscales separately for each sample based on the theoretical structure of the seven-factor model. The replicability of the factors in the inventory was evaluated by orthogonal Procrustes rotation. Factors, for which calculated coefficients have been found to be 0.80 or above, are virtually the same and tend to be judged as equal, while means and standard devia-

tions for the higher-order dimensions of temperament and character were calculated. The data from the three groups have been compared by ANOVA and t-tests for dependent samples. The eigenvalues of the character factors of the three data sets all suggested three factor solutions as adequate. Direct comparisons of factor structures of the TCI between the different language versions by means of orthogonal Procrustes rotations showed that the factor congruence coefficients for all comparisons were above 0.80 except those for the Persistence factor between the Swedish and the Iranians and between the Swedish and the refugees in Sweden for Reward Dependence. The results showed a relatively high degree of similarity in the factor structure of temperament and character in the three samples[23], but the differences between the Swedish and the Iranian individuals were (on average) larger, compared with those of the refugees. In this type of approach, the researcher try to overcome the influence of the cultural bias, checking both the validity and reliability of certain construct or its dimensions. But, following such an analysis, we are not able to know the exact level of the likelihood between the same constructs in different countries. This approach is more exact than first two in overcoming construct bias.

(d) Comparing »unframed« factor structure can be described as comparing the complete constructs with their best metric characteristics in relation to their best psychometric properties and the possibility of interpreting (it is best suited to certain cultures, and applying explorative strategy of factor analysis): In general, considering all the criteria for extracting and rotating factors in the factor analysis, we can classify these criteria on metric (like reliability, variance explained, the level of factor saturation, etc.) and meaning (interpretation) dimensions. In other words, the logic of extracting factors in the factor analysis can have as a result metrically perfect factors (but without logical sense) or interpretable and practically very applicable concept (but without reliability and construct validity). In a cross-cultural context, for each culture separately, we first have to find a metrically satisfying factor solution (valid and reliable), omitting the items that are not reliable or which low saturate the factor(s), in the next iterations of the factor analysis. When the metric solution is sufficient, the next step is checking the meaning criteria, again in each country (culture) separately. If the interpretation criteria are not satisfied, we can omit items which do not belong to the expected factor in the next iteration of the factor analysis. Finally, we stop with the iterations of the factor analysis when we found reliable, valid and interpretable factors, for all the items remained in the measuring instrument, for each culture separately. However, this type of comparison could be probably limited in application of variety of the data analysis methods, but can enable the interpretation of the aspects of construct bias, as well as the item bias.

(e) Checking the similarity of the constructs in the samples from different cultures (using structural equation modeling approach): Analyzing the likelihood that the constructs in the samples from different cultures are equal (using structural equation modeling approach) is the method that tries to overcome the deficiencies of the previous one. Structural equation modeling (SEM) approach is used to test for multigroup equivalence and to elucidate the many complexities that contribute to findings of nonequivalence across cultural groups. This advanced and rigorous approach to the test adaptation consists of three steps[1]: (a) translation of the initial instrument into the desired language using a combination of a team and back translation approach; (b) based on the hypothesized factorial structure of the original instrument, test for the validity and reliability of scores on the newly translated version based on representative samples within the target country; and (c) test the adapted instrument within its target country measurement and the structural equivalence across the target and source countries. These three global procedures can be much more complex, with many additional tests before becoming sufficiently equivalent to its parent version. Byrne[1] points out the two main concerns about the instrument's construct validity. The first is the extent to which items on the instrument are operating similarly across samples representative of each population (is the item content being interpreted and perceived in exactly the same way, or does the item measure the underlying construct to the same extent in both groups, or if the item format is operating equally well for both groups?). The second concern focuses on the underlying theoretical structure of the instrument (if the construct underlying each sub-dimension have the same meaning, if each has the same dimensional structure, and if relations among these constructs are group-equivalent). These two concerns have a focus only on the extent to which a measuring instrument is equivalent across groups, and have outlined particular cautions in the special case where an instrument is developed in one culture and then imported for use in another culture. The special use of SEM for analyzing multigroup equivalence is the framework of a confirmatory factor analytic (CFA) model. Few authors describe in detail the tests they use for multigroup equivalence[24,25]. Vandenberg and Lance[26] used diverse models based on the LISREL, AMOS, and EQS programs. In the research conducted by Bodkin-Andrews, O'Rourke, & Craven[27], using CFA techniques, testing of factorial invariance consisted of five increasingly restrictive models. The first is the least restrictive model (completely free), with no between-group invariance constraints placed on the estimated parameters. In the second model, the factor loadings were held invariant across the specified groups, which is usual the minimum condition of factorial invariance[28]. The third model held the factor loadings, factor variances and the covariances constant, and is often recommended as the minimal requirement of equivalence test, especially when dealing with more sensitive cultural groups[27]. The fourth model is based on holding the factor loadings and the uniquenesses invariant. The fifth model was the most restrictive in that it held all parameters invariant across the groups (totally invariant model). However, the process of testing for equivalence

can involve several additional steps depending on the aim of the study, the type of the data, and the level of stringency a researcher wishes to apply. Here are the two examples.

Horng and Teng[29] studied cross-cultural comparisons in quality measurement for undergraduate hospitality, tourism and leisure programmes. This study provides evidence for cross-validation of the previously constructed instrument, developed by Horng, Teng and Baum[30], based on applying confirmatory factor analysis (CFA) to a sample of faculty members from Taiwan and the USA. Multi-group confirmatory factor analysis indicated some support for the proposed cross-cultural invariance. CFA results satisfied some but not all requirements for cross--cultural equivalence. This evaluation represent the attempt to assess and report the cross-cultural validity of instruments applying multi-group SEM in hospitality and tourism education. Although the factor structure of the second-order model was confirmed and consistently displayed, some measurement items are the likely source of the variation in factor loading invariance among groups. The difference in measurement equivalence implies a lack of consensus about the operationalization of certain constructs representing quality in hospitality, tourism and leisure programmes across national boarders. The findings revealed that the cultural context needs to be considered in cross-cultural quality measurement.

Byrne & Watkins[27] had two purposes in conducting their study: (a) to test for the equivalence of a well--known measurement instruments across two culturally diverse groups and (b) based on the findings of nonequivalent items, to identify possible determinants of their noninvariance. The findings point to major differences between Australian and Nigerian adolescents with respect to self-perceived physical ability and self-perceived relations with parents (using Self Description Questionnaire I). Simultaneously, the research revealed similarly specified and well-fitting factor structures for both cultural groups, but also the evidence of both measurement and structural noninvariance. These results corrected previous interpretations regarding the equivalence of the instruments and raise a concern that measurement instruments (based on questionnaires) rarely can ever be totally equivalent, when used in cross-cultural comparisons.

The importance of using data analysis strategies for reducing the influence of the bias in cross-cultural research is not only restricted to cross-cultural psychological issues. In general, in each field of applied psychology (as well in medical anthropology) we can analyze psychological constructs related to health, such as perfectionism and hardiness. Perfectionism represents tendency to reach very high standards. Perfectionism is a set of cognitions, including expectations and interpretations of events and evaluation of self and others characterized by taking stands with a series of unrealistic standards, rigid and inflexible, that equal self-evaluation with success[31]. Perfectionism is often observed as neurotic disposition, associated with many psychopathological attributes: depression, feeding disorders[32]; social phobia/anxiety and obsessive-compulsive disorders[33]; with the feelings of loss and anxiety, guilt, delaying tasks, suicidal ideas, low self-esteem. The concept of hardiness has been used in an effort to explain different abilities of humans to face stress[34]. Hardiness explains why some individuals develop somatic and psychological illnesses when faced with stressful life situations, while the others remain »healthy«. The construct of hardiness consists of three positively intercorrelated, but not identical elements, the so called »three Cs«: commitment, control and challenge[35]. However, these constructs have the different metric expression in different cultures and have to be adjusted to a certain culture. In other words, the construct, method and item bias have to be overcome and neutralized. Reliability analysis and factor analysis techniques can deal simultaneously with all three types of bias.

Conclusion

In cross-cultural research, bias can be an indicator that instrument scores based on the same items measure different traits and characteristics across different cultural groups. To reduce construct, method and item bias, the researcher can draw upon at least a few strategies: simply comparing average results in certain measurement instruments; comparing only the reliability of certain dimensions of the measurement instruments, applied on the »target« and »source« samples of participants (from different cultures); comparing the »framed« factor structure of the measuring instruments, applied on the samples from the »target« and »source« cultures; using explorative factor analysis strategy on separate samples; comparing the complete constructs (»unframed« factor analysis) in relation to their best psychometric properties and the possibility of interpreting (best suited to certain cultures, applying explorative strategy of factor analysis); and checking the similarity of the constructs in the samples from different cultures (using structural equation modeling approach). Each approach has its advantages and disadvantages. However, the structural equation modeling approach is from a methodological perspective the fullest and the most appropriate approach for resolving all types of bias.

Acknowledgements

This article was supported by the Ministry of Science, Education and Sports of the Republic of Croatia (Grant No. 196-1962766-2751 to P.R., project »The population structure of Croatia – anthropogenetic approach«). I would like to thank dr. Linda A. Bennett and Josip Lah, B.Sc. for their valuable comments and help in translation.

REFERENCES

1. BYRNE BM, Testing for instrument equivalence in cross-cultural comparisons of self-concept: a look at aome of the complexities, accessed 30.9.2011. Available from: URL: http://www.self.ox.ac.uk/documents/Byrne.pdf — 2. VAN DEN BOS GR, APA dictionary of psychology (American psychological Association, Washington, DC, 2007). — 3. VAN DE VIJVER FJR, LEUNG K, J Pers, 69 (2001) 1007. — 4. VAN DE VIJVER F, TANZER NK, Eur Rev Appl Psychol, 54 (2004) 119. — 5. HUI CH, TRIANDIS HC, J Cross Cult Psychol, 20 (1989) 269. — 6. TRIANDIS HC, Am Psychol, 51 (1996) 407. — 7. MARIN G, GAMBA RJ, MARIN BV, J Cross Cult Psychol, 23 (1992) 498. — 8. GARTSTEIN MA, SLOBODSKAYA HR, ¯YLICZ PO, GOSZTYLA D, NAKAGAWA A, Rev Int Psicol Ter Psicol, 10 (2010) 55. — 9. CHEN Y, TISAK M, TISAK J, GOLDSTEIN S, Perspectives, 7 (2006) 47. — 10. CHRISTIE PMJ, KWON IWG, STOEBERL PA, BAUMHART, J Bus Ethics, 46 (2003) 263. — 11. BRISLIN RW, Int J Psychol, 11 (1976) 215. — 12. BRISLIN RW, J Cross Cult Psychol, 1 (1970) 185. — 13. HOFSTEDE G, Culture's consequences: comparing values, behaviors, institutions, and organizations across nations (2nd ed.) (Sage, Thousand Oaks, CA, 2001). — 14. CARRMINES EG, ZELLER RA, Reliability and validity assessment (Sage, Beverly Hills, CA, 1979). — 15. HEINE SJ, LEHMAN DR, GREENHOLTZ KP, J Pers Soc Psychol, 82 (2002) 903. — 16. PENG K, NISBETT RE, WONG NYC, Psychol Methods, 2 (1997) 329. — 17. TRIANDIS HC, MCCUSKER C, HUI H, J Pers Soc Psychol, 59 (1990) 1006. — 18. MERENDA PF, Cross-cultural adaptation of educational and psychological testing. In: HAMBLETON RK, MERENDA PF & SPIELBERGER CD (Eds) Adapting educational and psychological tests for cross-cultural assessment (Erlbaum, Mahwah, NJ, 2005). — 19. HAMBLETON RK, DE JONG J, Language Testing, 20 (2003) 127. — 20. BARRETT PM, SONDEREGER R, SONDEREGER NL, Behaviour Change, 19 (2002) 220. — 21. FORTMANN K, LESLIE C, CUNNINGHAM M, International Journal of Selection and Assessment, 10 (2002) 98. — 22. RICHTER J, BRÄNDSTRÖM S, EMAMI H, GHAZONOUR M, Coll Antropol, 28 (2004) 865. — 23. STALLINGS MC, HEWITT JK, CLONINGER CR, HEATH AC, EAVES LC, J Pers Soc Psychology, 70 (1996) 127. — 24. HORN JL, MCARDLE JJ, Exp Aging Res, 18 (1992) 117. — 25. BYRNE BM, WATKINS D, J Cross Cult Psychol, 34 (2003) 155. — 26. VANDENBERG RJ, LANCE CE, Organizational Research Methods, 3 (2000) 4. — 27. BODKIN-ANDREWS G, O'ROURKE V, CRAVEN RG, Australian Journal of Education, 54 (2010), accessed 30.9.2011. Available from: URL: http://findarticles.com/p/articles/mi_hb6475/is_3_54/ai_n56-380967/?tag=content;col1. — 28. BYRNE BM, Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming (Erlbaum, Mahwah, NJ, 1998). — 29. HORNG JS, TENG CC, Journal of Hospitality, Leisure, Sport and Tourism Education, 10 (2011) 49. — 30. HORNG JS, TENG CC, BAUM T, Journal of Hospitality, Leisure, Sport and Tourism Education, 8 (2009) 4. — 31. HEWITT PL, FLETT GL, J Pers Soc Psychol, 60 (1991) 456. — 32. IVANOV L, PENEZIĆ Z, Burns Perfectionism Scale. In: PROROKOVIĆ A, LACKOVIĆ-GRGIN, K, ĆUBELA V, PENEZIĆ Z. (Eds) In Croat. Collection of psychological scales and questionnaires 2 (2004) 13. — 33. MARTIN KA, BRAWLEY LR, Pers Individ Dif, 27 (1999) 901. — 34. KOBASA SC, J Pers Soc Psychol, 37 (1979) 1. — 35. 8. MADDI SR, HESS MJ, Sport Psychology, 23 (1992) 360.

J. Sindik

Institute for Anthropology Research, Gajeva 32, 10000 Zagreb, Hrvatska
e-mail: josko.sindik@inantro.hr

STRATEGIJE OBRADE PODATAKA ZA SMANJENJE UTJECAJA PRISTRANOSTI U MEĐUKULTURALNIM ISTRAŽIVANJIMA

SAŽETAK

U međukulturnim istraživanjima, znanstvenici moraju prilagoditi i konstrukte i pripadne mjerne instrumente, koji su razvijeni u jednoj kulturi, a zatim uvezene za uporabu u drugoj kulturi. Uvoz pojmova iz drugih kultura često se pojednostavljeno svodi isključivo na jezično podešavanje mjernih instrumenata koji definiraju određeni (psihološki) konstrukt U kontekstu međukulturnih istraživanja, testna pristranost može se definirati kao opći pojam za sve čimbenike koji su smetnja valjanosti međukulturalne usporedbe. Pristranost može biti pokazatelj da rezultati u nekom mjernom instrumenta, na temelju istih čestica, mjere različite crte i karakteristike ljudi iz različitih kulturnih grupa. Za smanjenje konstruktne, metodološke i čestične pristranosti, istraživač se može koristiti s bar nekoliko strategija: (1) jednostavnom usporedbom prosječnih rezultata u određenim mjernim instrumentima; (2) uspoređujući samo pouzdanosti određenih dimenzija mjernih instrumenata primijenjenih u ciljnim i izvornim uzorcima sudionika (iz različitih kultura); (3) uspoređujući »uokvirenu« faktorsku strukturu (fiksirani broj faktora) mjernih instrumenata, primijenjenih na uzorcima iz ciljnih i izvornih kultura, korištenjem eksploratorne strategije faktorske analize na svakom uzorku zasebno; (4) usporedbom kompletnih konstrukata (faktorska analiza »bez okvira«, tj. neograničeni broj faktora) po kriteriju napogodnijih metrijskih karakteristika ali i kriteriju interpretabilnosti (najbolje prilagođenih određenim kulturama, uz korištenje eksploratorne strategije faktorske analize); (5) provjera sličnosti konstrukata na uzorcima iz različitih kultura (korištenjem pristupa modeliranja strukturalnim jednadžbama). Raspravljene su i prednosti i nedostaci svakog od opisanih pristupa.